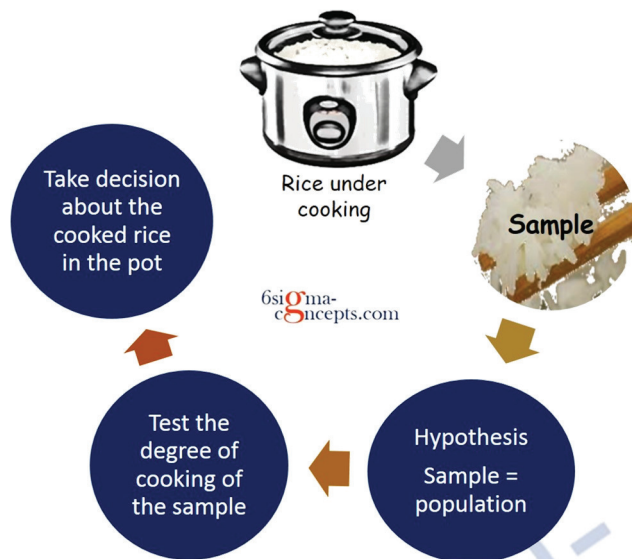


Rice Cooking and the Inferential Statistics

Every one of us must have cooked the rice in kitchens. What actually we do is to draw a sample of 4-5 grains of rice from the pot and press it between the fingers and decide whether the rice in the cooking pot is cooked or not. *Isn't it surprising? Based on the property of few rice grains we are taking decision on whole pot of rice! – That's the inferential statistics.*



Same hypothesis testing is going in your mind when you are buying 5 Kg of tomatoes based on a sample of 5-6 tomatoes that you tested before you took 5 Kg of it. After reaching homes you found that some of them have to be thrown immediately.

There will always be some degree of uncertainty whenever we are using a sample to predict the property of the population.

What we are doing unknowingly is that we **develop a hypothesis in our mind** that the degree of cooking of the sample is equal to the degree of cooked rice in the pot, this is called as **null hypothesis**. Then we draw a sample from the pot for **testing the hypothesis**. There can be two possible outcomes,

1. Rice is well cooked, in that case we accept the null hypothesis.
2. Rice is not well cooked (it can be either undercooked or over cooked). This becomes the alternate hypothesis


Now we subject the sample of rice to test by pressing it between our fingers. This is called as **test statistics**. Based on the test statistics we can take following two decisions

1. The null hypothesis can't be rejected (we accept that the rice in the pot is well cooked).
2. There is not enough evidence to accept the null hypothesis (the rice in the pot is either overcooked or under cooked).

You might be wondering that why we can't directly take decision to accept or reject the null hypothesis. Why we are making above indirect statement about the null hypothesis?

This is because there will always be some uncertainty in taking the correct decision. This is because *hypothesis testing is performed for predicting the characteristics of the population based on the sample information* and we must allow for the possibility of errors because of sampling error. These errors may occur because of sampling issues (e.g. sample was taken from the

top of the pot, but at bottom the rice was burnt). There can be two kinds of errors that can be made in above hypothesis testing.

		Population parameter	
		Null hypothesis is true	Alternate hypothesis is true
Conclusion	Accept the null hypothesis	Correct Decision	Type-II error: Accepting the null hypothesis when actually it was false <i>Even though the rice in the pot was not well cooked we declared that it is based on the sample statistics.</i>
	Reject null hypothesis	Type-I error: Rejecting the null hypothesis when actually it was true. <i>Even though the rice in the pot was well cooked we declared that it is not based on the sample statistics.</i>	Correct Decision 

In above examples we have access to the population whether it is a pot of rice or the 5 Kg of tomatoes. Imagine the scenarios where we don't have access to the population's information.

1. Your office requires 500 diaries for the employees, based on the sample of 1-2 diaries provided by some of the vendors, order is released to one of them.
2. You are in production planning and need to place an order for a raw material for the entire month of the production. The QC department approved 2-3 test samples which your vendor has provided to you. Based on this 2-3 samples you released the purchase order for the entire month to that vendor.
3. A lot of 50000 LED has reached your warehouse to be used latter in one of the electronic gadget. Have you ever wondered that by testing few of those LED by the QC would result in acceptance or rejection of the whole lot!

Statistics provide us a handle that enables us to take decisions in presence of uncertainty.

Hence in order to deduce inferences from data one should take into account not only the mean, but also the variance embedded in the system. Six-sigma help us in reducing known source of variations thereby reducing the margin of error so that we come close to the population parameters.